# A Single-cell and Spatial RNA-seq Database for Alzheimer's Disease (ssREAD)

Cankun Wang[1], Diana Acosta[2], Megan McNutt[1], Jiang Bian[3], Anjun Ma[1], Hongjun Fu[2,4*], Qin Ma[1,*]

[1] Department of Biomedical Informatics, The Ohio State University, OH 43210, USA

[2] Department of Neuroscience, The Ohio State University, OH 43210, USA

[3] Department of Health Outcomes & Biomedical Informatics, University of Florida, FL 32606, USA

[4] Chronic Brain Injury Program, The Ohio State University, OH 43210, USA

*Correspondence: hongjun.fu@osumc.edu and qin.ma@osumc.edu

## Abstract

Alzheimer's Disease (AD) pathology has been increasingly explored through single-cell and single-nucleus RNA-sequencing (scRNA-seq & snRNA-seq) and spatial transcriptomics (ST). However, the surge in data demands a comprehensive, user-friendly repository. Addressing this, we introduce a single-cell and spatial RNA-seq database for Alzheimer's disease (ssREAD). It offers a broader spectrum of AD-related datasets, an optimized analytical pipeline, and improved usability. The database encompasses 1,053 samples (277 integrated datasets) from 67 AD-related scRNA-seq & snRNA-seq studies, totaling 7,332,202 cells. Additionally, it archives 381 ST datasets from 18 human and mouse brain studies. Each dataset is annotated with details such as species, gender, brain region, disease/control status, age, and AD Braak stages. ssREAD also provides an analysis suite for cell clustering, identification of differentially expressed and spatially variable genes, cell-type-specific marker genes and regulons, and spot deconvolution for integrative analysis. ssREAD is freely available at https://bmblx.bmi.osumc.edu/ssread/.

**Keywords:** Alzheimer's Disease, Web database, single-cell and single-nucleus RNA sequencing, Spatial transcriptomics

## Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder, the most common form of dementia worldwide. Over 57 million individuals globally suffer from this debilitating condition[1]. Despite the significant strides in medical research and development, therapeutic interventions for AD remain distressingly ineffective. This glaring lacuna in clinical therapeutics underscores the urgency to probe into the intricate molecular mechanisms underlying the disease's cellular and regional susceptibility that are still largely enigmatic[2]. Recent high-throughput sequencing technologies, particularly single-cell RNA-sequencing (scRNA-seq) and single-nucleus RNA-sequencing (snRNA-seq) have cast fresh light on our exploration of AD pathogenesis. To study the cellular heterogeneity of the brain and reveal the complex cellular changes in AD, we

launched scREAD in 2020[3]. By then, it was the first database dedicated to managing public AD-related sc/snRNA-Seq data from human and mouse brain tissue. As the sequencing technology and scientific inquiry rapidly evolved, more studies are discerning the spatial information of differentially expressed genes (DEGs) associated with AD pathology, the interconnectedness of DEGs related to AD biomarkers, DEGs enriched in specific cell subtypes, cell-cell communications, and regional and cellular vulnerability in AD[4]. Spatial transcriptomics (ST) revolutionized our understanding of neurobiology and AD pathogenesis by enabling the visualization of gene expression patterns within their spatial context. Yet, these public-available scRNA-seq, snRNA-seq, and ST data have not been well collected and managed by any AD databases. In addition, the remarkable potential usage of these datasets is accompanied by the formidable challenge of data aggregation, analysis, and interpretation, necessitating substantial computational resources and bioinformatics expertise.

In recent years, there has been a significant increase in the availability of AD-related sc/sn RNA-seq datasets. This has led to the development of various databases, including those serve as general repositories for scRNA-seq data and others that specifically focus on AD and the brain. For instance, the TACA[5] database facilitates differential expression comparisons to identify cell type-specific gene expression alterations, cell-cell interactions, and drug screening opportunities. Meanwhile, the SC2Disease database[6] aims to offer a comprehensive and accurate resource of gene expression profiles across various cell types for 25 diseases. However, the AD data provided by SC2Disease is limited to a single dataset from one brain region (prefrontal cortex). SCAD-Brain[7] is another public database dedicated to AD, with a focus on both human and mouse brain data. It provides more extensive analysis results, such as cell communication analysis and trajectory information. Despite these advances, there remains a gap in the field for a specialized database that concentrates on spatial transcriptomics in AD and offers comprehensive differential analyses under various conditions, such as sex-specific, region-specific, and comparisons between AD and control groups. Integrating these diverse datasets and conditions would prove invaluable for researchers studying the complex landscape of AD'.

To address these burgeoning complexities and to meet the scientific community's growing demand for comprehensive, integrated, and accessible data analysis, we introduce **ssREAD** (Single-cell and Spatial RNA-seq databasE for Alzheimer's Disease). It includes 381 ST and 277 sc/snRNA-seq AD-related datasets. These sequencing data enable researchers to investigate transcriptomic alterations in AD compared to the control and their regulatory mechanisms at various resolutions: sub-cellular, cellular, and spatial levels[8,9], which will help uncover the pathogenesis of AD. We also highlight the sc/snRNA-seq and ST data analysis framework in ssREAD, including cell clustering and annotation, DEGs and spatially variable gene identification, cell-type-specific regulon

inference, cell-cell communication analysis and functional enrichment analysis. Moreover, the integrative exploration of ST and sc/snRNA-seq data revealed nuanced molecular landscapes that underlie AD, emphasizing ssREAD's unparalleled capability. Beyond that, ssREAD also delivers marked improvements to the user interface. These modifications, grounded in user-centric design principles, advance visibility and usability, fostering an environment conducive to intuitive data visualization and streamlined querying.

## Results

### 1. Overview of ssREAD

ssREAD comprises 381 ST samples from 16 AD-related studies and 1,053 sc/snRNA-seq samples from 85 studies (**Supplementary Data 1**). The 1,053 sc/snRNA-seq samples are grouped into 277 datasets by integrating sample replications from the same study. All datasets are collected and downloaded from Broad Institute SingleCellPortal, Gene Expression Omnibus (GEO), and Synapse[10] (**Supplementary Fig. 1**). ssREAD has a considerable 379% increase in the number of integrated sc/snRNA-seq datasets than the previous scREAD (from 73 to 277), encompassing over three times the total cell and nucleus count.

In our pursuit of thoroughness, each dataset is meticulously annotated, providing pertinent details such as species, gender, brain region, disease/control distinction, and AD Braak stages. In the realm of species, 144 out of 277 integrated sc/snRNA-seq datasets are from human samples, whereas 133 datasets are from mouse samples. In contrast, mouse datasets are prominently represented with 319 ST datasets, four times more than human ST datasets **(Fig. 1A).** When demarcating datasets based on the AD condition, 154 integrated sc/snRNA-seq datasets are assigned to the AD group, as opposed to the 123 datasets associated with control scenarios (**Fig. 1B**). This distribution takes a nuanced shift within ST samples, with AD-oriented datasets numbering 140 and control datasets as 241. Sex-specific stratification is also evident within the ssREAD construct. It includes 140 integrated datasets from sc/snRNA-seq and an additional 213 datasets sourced from ST methodologies (**Fig. 1C**), compared to 137 sc/snRNA-seq and 168 ST datasets for females. ssREAD also offers a detailed breakdown of samples by brain region, with separate datasets for sc/snRNA-seq and ST techniques (**Fig. 1D**). ssREAD showcases a range of technologies that contribute to a comprehensive spatial transcriptomics profile (**Fig. 1E**). For the mouse model, technologies span from Slide-seq (106 samples) to HDST (3 samples). The human segment is inclined towards 10x Visium, accounting for 18 samples, closely trailed by In-situ ST of 20 samples, and MERFISH with 24 samples.

Benchmarking ssREAD against contemporaneous repositories underscores its unparalleled stature. Compared with some existing databases with a collection of

datasets on AD, such as scREAD[3], SCAD-Brain[7], TACA[5], SODB[11], STOmicsDB[12], and Single-Cell Portal, ssREAD has the most decadent AD-related samples collected for both sc/snRNA-seq and ST data with great depth and breadth (**Fig. 1F**). Overall, ssREAD leads to a comprehensive collection of AD-related scRNA-seq, snRNA-seq, and ST datasets, offering both breadth and depth of data that stand unrivaled compared to other available resources or databases.

Our ssREAD collects the above data and offers comprehensive, in-depth data analyses and result interpretations. For sc/snRNA-seq data, ssREAD provides functions including cell clustering, cell type annotation, marker gene expression visualization, and cell proportion analysis (**Fig. 2A**). For ST data, it provides visualizations for original spatial H&E image, layer/tissue architecture/spatial domain annotation, marker gene expression on spatial map, and spot deconvolution (**Fig. 2B**). DEGs can be identified for cell types in sc/snRNA-seq data or spatial layers in ST data (**Fig. 2C**). Cross-data queries and analyses are also facilitated, including comparative studies between male and female subjects across various datasets. Additionally, spatially variable genes (SVGs) can be identified via spaGCN[13] from ST data to show marker genes with spatially resolved expression patterns that may be related to tissue functions. Functional enrichment analysis is included to identify pathways or gene ontology enriched by DEGs or SVGs (**Fig. 2D**). Moreover, ssREAD also features cell-type-specific (or layer-specific) regulons for individual datasets and the integrated cell atlas (**Fig. 2E**), focusing on cellular and regional vulnerability in AD.

To ensure ease of use, all analytical results are displayed via a user-friendly and single-access web portal that frees AD researchers from the requirement of extensive programming knowledge. We offer new interactive plots for visualizing cells and spatial spots, including scatter plots, bar plots, and violin plots, as well as real-time analyses for DEGs, SVGs, and functional enrichment queries. Furthermore, all datasets, including author-provided metadata and cell type labels, are provided in ready-to-analyze formats (e.g., .h5ad and .h5seurat), compatible with analysis tools such as Seurat and Squidpy for further analysis.

**2. Spatially-informed Subpopulation Analysis Reveals Cellular Heterogeneity in AD**

To illustrate the ST data analysis workflow and functions that are available to users of ssREAD, we used two ST data (ST01101 and ST01103), labeled by six cortical layers and the adjacent white matter in two human middle temporal gyrus (MTG) brain samples based on the information provided in the original study[14] (**Fig. 3A**). We performed DEG analysis between AD (ST01103) and control (ST01101) for each functional layer (**Supplementary Data 2**) and evaluated the DEG consistency with the other two comparison groups (i.e., ST01102 vs ST01104 and ST01106 vs ST01105)

(**Supplementary Data 3-4**). Results showed that there exist DEGs between AD and control in each layer that overlapped among the three comparison groups (**Supplementary Fig. 2**). The overlapping DEGs among the three comparison groups correspond to up-regulated genes in AD compared to controls of layer 4 with three DEGs, layer 5 with two DEGs, and layer 6 with three DEGs. Further implementation of RESEPT[15], a deep-learning framework for spatial domain detection, provided a precise delineation of the tissue architecture and functional zones in both control and AD brain tissues (**Fig. 3B**). The resulting spatial delineation remains comparable to the six cortical layers identified in the original study. Layers 5-6 of the AD sample (ST01103) exhibit slight differences in their delineation compared to controls and their original labels, suggesting there may be underlying differences in the functional zones of these layers. The data is consistent with previous publications suggesting layer 5 is highly relevant to changes in AD including accumulation of neurofibrillary tau tangles[16-18]. Moreover, the potential of ssREAD in navigating the complex spatial information of AD was further exemplified through a multi-dimensional exploration of spatially informed sub-populations via MAPLE[19] (**Fig. 3C**). Noted that MAPLE has a critical multi-sample design considering information sharing across samples and accommodating spatial correlations in gene expression patterns within samples. Thus, clusters MAPLE identifies are sample-specific and could be either shared among samples (e.g., cluster 1 in both ST01101 and ST01103) or unique in individual samples (e.g., cluster 5 only in ST01103). This analysis demonstrated ability of ssREAD to untangle the spatial complexity inherent in the transcriptomic landscape of AD. To further illuminate these cellular dynamics, we mapped the MAPLE cluster annotations in individual samples to their corresponding layer annotations (**Fig. 3D, Supplementary Fig. 3, and Source Data 1**). In doing so, we show that cluster 1 corresponds to layer 5 and 6 of the AD sample (ST01103), but only corresponds to layer 6 of the control sample (ST01101). This indicates the deviations in layer 5 of AD cases compared to controls may also be a result of differences in their spatial organization as well as their gene expression patterns.

At the molecular level, we identified DEGs in each MAPLE cluster (**Fig. 3E and Source Data 2**) and between control and AD samples within individual MAPLE clusters (**Fig. 3F, Supplementary Data 5, and Source Data 3**). The expression pattern of these DEGs underscores the molecular heterogeneity in MAPLE clusters (subpopulations) and between AD and the control (**Fig. 3F**). For example, the expression of *DEPP1*, also known as PGC-1α (peroxisome proliferator-activated receptor gamma coactivator 1-alpha), was found to be significantly lower in AD than the control in our dataset. The PGC-1α is highly responsive to numerous forms of environmental stress, including

temperature and nutritional status. Several studies have reported that the level of PGC-1α significantly decreases in the brains with AD compared to control brains[20-23]. PGC-1α has thus been suggested to contribute to the improvement of AD pathophysiology. Additionally, the pathway enrichment analysis on AD and control DEGs in cluster 1 showed both upregulated and downregulated pathways. Interestingly, all five downregulated pathways are associated with immune responses/functions (**Fig. 3G**), which may highlight the important role of immune cells in AD pathogenesis[24-27]. We then further investigated the association of genes within each pathway to microglial states previously published by Sun *et al.* [28], as well as disease-associated microglia (DAM) and activated response microglia (ARM) genes[29,30]. In agreement with previously published data, three out of five of our upregulated pathways (Regulation of Expression of Slits and Robos, Selenoamino Acid Metabolism, and Eukaryotic Translation Elongation) include several genes that make up a population of microglia (MG3) that is highly enriched with disease associated microglial genes. On the other hand, the five downregulated pathways (Scavenging of Heme from Plasma, Interleukin 10 Signaling, Tnfs Bind their Physiological Receptors, Fcgr Activation, and Creation of C4 and C2 Activators) are composed of genes found in microglia states that are associated with inflammation due to presence of cytokine and cytokine receptor related genes (**Supplementary Fig. 4**). Therefore, the upregulated pathways overlap with disease associated microglia states while downregulated pathways do not include genes found in disease-associated microglia states. In line with previous findings, inflammatory states of microglia change in relation to AD progression, with more strongly correlated inflammatory states being present at early disease stages and weaker inflammatory states present at late disease stages. Therefore, the downregulation of inflammatory related states in our data coincides with the dysregulation of microglia states found in AD. Besides DEGs and DEG enriched pathways, we also identified 305 SVGs in cluster 5, such as *PLP1* and *UCHL1* (**Fig. 3H**), which show clear spatial expression patterns that are linked to specific tissue layers. Interestingly, both genes have also been found to be associated with AD[31,32].

To further investigate the regulatory mechanisms in different clusters between the above Control (ST01101) and AD (ST01103) samples, we implement DeepMAPS[33] in the ssREAD framework for spatial transcriptomics-guided gene regulatory network analysis. This facilitated a comprehensive investigation into the interconnected relationships among key transcription factors (TFs) and their associated genes. Our analysis highlighted a network comprising 10 TFs (i.e., *NR1H3*, *SREBF1*, *ATF6*, *TAL1*, *SOX10*, *NFYA*, *AR*, *MYC*, *HIF1A*, and *EGR1*) and 2,164 genes regulated by those TFs in MAPLE cluster 1 (**Supplementary Data 6**). This network, interconnected with the enriched genes from gene modules identified by DeepMAPS, underscores the dynamic and intricate interactions between critical TFs and their downstream regulated genes. Further dissection of these TFs showed differences in TF activities between control and AD samples, for example, *ATF6* and *EGR1* (**Fig. 3I**). These two regulators are reported to be

associated with AD pathology and highly related to stress response. *ATF6*, a key player in unfolded protein response to endoplasmic reticulum (ER) stress, has been found to reduce amyloid-beta (Aβ) toxicity via the downregulation of β-site APP-cleaving enzyme 1 (BACE1)[34]. On the other hand, *EGR1* may impair the brain's cholinergic function in the preclinical stages of AD via the upregulation of acetylcholinesterase (AChE)[35], and *EGR1* regulates tau phosphorylation and Aβ synthesis in the brain by enhancing activities of Cdk5 and BACE-1, respectively[36]. The GO and REACTOME enrichment analysis showed that genes regulated by these two TFs are enriched in multiple stress-related pathways. Our results showed that genes, including *XBP1*, *HSPA5*, *DDIT3*, *SEL1L*, *ATP2A2*, and *HSP90B1*, regulated by *ATF6* are related to response to ER stress pathways (**Supplementary Data 7**). While 106 genes regulated by *EGR1* are related to stress response, such as *CDKN1A*, *MYC*, *TP53*, and *RXRA*, responding to pathways including oxidative stress, ER stress, chemical stress, and stress-activated MAPK cascade. Overall, our data-driven analysis of ST data provides a steppingstone for future studies aimed at deciphering the complex molecular pathogenesis and novel therapeutic targets.

## 3. ssREAD Unveils AD Pathophysiology through an Integrated Analysis of Spatial and Single-cell Transcriptomics

One of the most prominent analyses that ssREAD can power is the spot deconvolution enabled by the integration of sc/snRNA-seq and ST data. A cornerstone in ssREAD is the utilization of the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) with ID of AD03501, an atlas that includes cells derived from the health and AD human middle temporal gyrus (MTG)[37]. The SEA-AD atlas originally encompassed an extensive dataset of 378,211 cells. We employed the sketch-based analysis feature in Seurat v5 to streamline our analysis, which strategically selects 'subset' or 'sketch' of 50,000 cells. This analytical decision was driven by achieving computational efficiency while ensuring a robust representation of cellular diversity. Such an atlas painted a detailed cellular tableau, effectively revealing cell distributions of 23 cell types (**Fig. 4A)**. Beyond cell types, the atlas also includes a nuanced statistical breakdown of the cells, distinctly categorized by the Braak stage, Thal phase, gender, and ethnicity, offering a comprehensive glimpse into the atlas's cellular constitution **(Fig. 4B-F)**. As shown, there is no batch effect among samples regarding Braak stage, Thal phase, and ethnicity. However, there are obvious sex-oriented differences in cell clusters, which may contribute to the possible pathological sex-bias differences in AD.

DEG analysis between AD and control cells revealed distinctive signatures in each cell type (**Supplementary Data 8**). For example, DEGs associated with homeostatic state of microglia (e.g., *P2RY12*, *CSF1R*, *CX3XR1*, *TGFBR1*, *MEF2A*, and *ENTPD1*) are decreased, while DEGs associated with the dyshomeostatic state of microglia (e.g., *CTSD*, *APOE*, *AXL*, *SPP1*, and *GPNMB*) are increased in AD compared to the control[30,38]

(**Fig. 4G**). To showcase the feasibility and power of our analysis, we further included a large AD dataset from the prefrontal cortex (PFC) region[39], and integrated the MTG sample with the PFC samples. The top 25 microglia DEGs between AD and control in the integrated data include most genes found in the individual datasets (**Supplementary Data 9-11**). We also identified 68 upregulated Microglia DEGs between AD and control overlapped across all datasets (**Fig. 4H, Supplementary Data 12, and Source Data 4**), indicating many DEGs can be recapitulated by using the integrated datasets. Importantly, our comparison highlights the differences in DEGs due to region differences, in which the MTG encompasses many identified DEGs compared to the PFC. This showcases the underlying transcriptomic changes may be more prominent in regions that are affected earlier in Alzheimer's disease in comparison to regions such as the PFC which are affected later in disease. In addition, the integrated data set reveals 17 DEGs that are not present in either individual dataset. The 17 DEGs suggest there are potential transcriptomic changes that are independent of region-based characterization which could only be identified by integrating such datasets. These genes include *DOPEY2*, *CSF2RA*, and SRRM2 which have been previously linked to AD[40]. For example, *CSF2RA* was upregulated in mouse microglia that have internalized Aβ plaques, and have been treated with IL-1β[40]. *DOPEY* was previously identified in cases (AD and/or MCI) in the ADNI and NIA-LOAD/NCRAD Family studies[41]. Lastly, *SRRM2* was previously studied due to its association with the progression of tauopathy in transgenic mice[42] as well as recruitment to tau aggregates[43]. We compared our DEG results from the integrated sc/snRNA-seq dataset to previously published spatial datasets[9]. Out of the 68 genes that are upregulated in the microglia population, we identified 29 genes (~43%) overlapped with DEGs from previously published spatial transcriptomics datasets (**Supplementary Data 13**).

Bridging the understanding between single-cell and spatial data, ssREAD was instrumental in performing a cell-type deconvolution analysis between the ST datasets and the SEA-AD atlas using the CARD R package[44]. This synergetic approach underscored the presence and dynamics of critical cell types, including oligodendrocytes, astrocytes, and endothelial cells, across both AD (ST01103) and the control (ST01101) datasets. The cellular insight was further illuminated by visualizing the expression of marker genes, such as *GFAP* (**Fig. 4I**) for astrocytes (**Fig. 4J**) and *MOBP* (**Fig. 4K**) for oligodendrocytes (**Fig. 4L**), distinguishing between control and AD conditions. Unfortunately, we observed a very low proportion of the microglia cell type in both Visium samples (**Supplementary Fig. 3B**). We further compared DEGs between AD and control in MAPLE cluster 1 to the identified single-cell DEGs from the integrated AD035 datasets. Among the overlapped DEGs, we highlight two upregulated (*GJA1* and *MT-ATP8*) (**Fig. 4M-N**) and two downregulated DEGs (*IFITM3* and *TUBB2B*) (**Fig. 4Q-R**) in the astrocyte population and their spatial distribution. Importantly, *GJA1* has been identified as a key regulator in AD pathogenesis, being associated with AD amyloid, tau pathology, and

cognitive functions. While depletion of astrocytic *GJA1* is linked to neuroprotection in neurons[45]. In Oligodendrocytes, we highlight four upregulated DEGs, *ERBIN*, *GPRC5B*, *MID1IP1*, and *SLC44A1* (**Figures 4O, P, S, T**). Notably, *ERBIN* and *MID1IP1* have been identified in previous datasets as upregulated in Oligodendrocytes for AD pathology cases[46]. The seamless fusion of snRNA-seq and spatial transcriptomics data, as facilitated by ssREAD, positions this integrated methodology as a potent tool for in-depth cellular analysis, demonstrating the platform's pivotal role in multi-dimensional data integration and offering transformative insights into AD[45]. As a result, we showcase the capability of ssREAD in integrating sc/snRNA-seq and ST data with computational tools like MAPLE, shedding light on the multi-dimensional dynamics of spatially informed subpopulations in AD. By unveiling the biological pathways and regulatory networks associated with disease progression, ssREAD allows a comprehensive understanding of the cellular and molecular landscape in AD, thus bringing us a step closer to unraveling the mysteries of this complex neurodegenerative disease.

## 4. Unveiling Sex-Specific Differences in Alzheimer's Disease at the Cellular Level

Besides the data-driven analysis that can be elaborated from ssREAD, we also showcase the ability to generate biological hypotheses that can be powered by the database. Capitalizing on the wealth of information offered by sc/snRNA-seq data, our investigation delves into the intricate, sex-specific nuances of AD at the cellular level. Using ssREAD, we were able to elucidate the AD heterogeneity between male and female, an aspect not previously pursued in the original research, which primarily focused on the molecular landscape of the over 183k cells in human brain hippocampus vasculature in AD[47] (AD019). Our investigation utilized all 16 samples from the original study (**Supplementary Data 14**). Commencing with a UMAP visualization of the single-cell data, color-coded by cell type (**Fig. 5A**), we achieved a broad perspective of 13 cell types (i.e., Arterial cell, Astrocyte, Capillary cell, Ependymal cell, Fibroblast, Microglia, Neuron, Oligodendrocyte progenitor cell, Oligodendrocytes, Pericyte, Smooth muscle cell, T cell, and Venous). From the ensuing breakdown of the proportion and count of each cell type by sex (**Fig. 5B-C**), we observe more Oligodendrocytes, Astrocytes, and Microglia in overall female cell types than male. Such sex-specific differences in cellular composition raise intriguing questions about the potential roles of these variations in disease pathogenesis and progression, warranting further investigation. We further compared DEGs across four demographic groups: Male AD patients, Female AD patients, Male controls, and Female controls. The result showcased both unique and shared gene signatures among these groups (**Fig. 5D**). Our findings indicate that both sex and disease status can shape the transcriptomic landscape of cells, which could have profound implications for understanding the underlying molecular mechanisms of AD and developing targeted therapies.

By comparing the gene expressions between males and females, we revealed unique sex-specific gene signatures within different cell types (**Supplementary Data 15**). Specifically, we identified 44 upregulated and 108 downregulated genes in male microglia vs female (**Fig. 5E**). We highlight the top 10 upregulated and downregulated DEGs, in which we identify *ATP1B3*, a previously identified ARM and DAM gene is downregulated in our comparison of female microglia to male microglia[29,30]. On the other hand, *LHFPL2*, and *RTTN* are DAM genes that we find are upregulated in female microglia compared to male[30]. Our data also remains consistent with a previous publication of AD-related transcriptional sex differences that identified *ACSL1*, *ATP1B3*, *CD163*, *CREM*, and *SRGN* as downregulated in female microglia cell nuclei compared to male[48]. The upregulation of genes *APBB1IP*, *GLDN*, *LHFPL2*, *LRRK1*, *P2RY12*, *RASSF8*, and *RTTN* in female microglia compared to male is also consistent with previous findings[48]. We validated the sex related DEGs findings using two additional datasets (ID AD035 and AD048) (**Supplementary Data 16-17** and **Supplementary Fig. 5**). We find that the Y chromosomal genes *UTY*, *LINC00278*, *TTTY14*, and *USP9Y* are found in both datasets and correspond to genes that are upregulated in male microglia; While the X chromosomal gene, XIST, is upregulated in female microglia and is found in both datasets. Our analysis of Gene Set Enrichment Analysis (GSEA) further explored the activity of binding and uptake of ligands by scavenger receptors, which are known to be involved in the recognition and clearance of various ligands, including modified lipoproteins, cellular debris, and pathogens[49]. In AD, the dysregulation of scavenger receptor activity has been implicated in the clearance of Aβ plaques (**Fig. 5G**), which are one of the hallmark pathological features of the disease[50]. Previous studies have reported sex differences in scavenger receptor expression in the context of AD[51]. Additionally, we observed differences in GSEA enrichment of neurogenesis pathways in Microglia under AD conditions. Microglia can shape adult hippocampal neurogenesis[52]. We compared Male AD vs. Male Control (**Fig. 5H**), Female AD vs. Female Control (**Fig. 5I**), and Male AD vs. Female AD (**Fig. 5J**). Interestingly, neurogenesis pathway activity varied significantly, with upregulation observed in female AD, and downregulation in male AD patients. This observation denotes a sex-dependent dysregulation of neurogenesis in AD, indicating that the disease could affect foundational neural processes in a sex-specific manner. Enhancing our understanding of how microglial activation states differentially regulate adult neurogenesis in men and women could yield invaluable insights into the disease's intricate mechanisms. Taken together, these results underscore the intricate interplay between sex and cellular and molecular profiles in AD. They highlight the necessity of considering sex as an integral factor in AD research, and point towards the potential for developing more personalized, sex-specific therapeutic strategies in the future.

## Discussion

We develop ssREAD, a single-cell and spatial RNA-seq database for AD, not only accommodates the expansion of sc/snRNA-seq data but is uniquely poised to incorporate and harness the emergent wealth of AD-related ST data. ssREAD houses data encompassing various species, diseases, tissues, and cell types, thereby permitting granular analyses that could reveal intricate biological phenomena. ssREAD's capabilities are broad, allowing for diverse analytical activities such as contrasting diseased tissues against healthy controls, identifying distinct cell types through gene markers, and examining gene expression across diverse tissues and cell types. The most compelling findings of sex differences in AD unveil a profound heterogeneity between male and female cellular profiles. Particularly, females exhibited a heightened presence of Oligodendrocytes, Astrocytes, and Microglia compared to males. Furthermore, while female AD patients showed upregulation in neurogenesis pathway activity, their male counterparts displayed a stark downregulation. These insights align with the findings of previous studies claiming that transcriptional responses were substantially different between sexes in different cell types[53], underscoring the pivotal role of gender in shaping AD's cellular and molecular features. Last but not least, designed with the end-user in mind, ssREAD is characterized by its user-friendly interface, query function, and visualizations. In terms of its infrastructure, ssREAD utilizes high-performance computing to manage large-scale single-cell data analysis efficiently. It comprises a cutting-edge server architecture employing diverse programming languages, machine learning frameworks, and data visualization libraries. As our understanding of the complexity and heterogeneity of biological systems continues to deepen, tools like ssREAD can play an increasingly vital role in our pursuit and understanding of AD studies.

To ensure the stability and availability of ssREAD, we have set up backup servers and have implemented cloud-based backup solutions. In case of unavailability of the main server, our development team can switch the link to redirect to an alternative server without requiring users to enter a different URL. To improve transparency and provide real-time updates for our users, we have created a status page for ssREAD (https://ssread.statuspage.io/). This page features automated monitoring of the server's status and allows users to track the operational status of the database, including any scheduled maintenance or unexpected downtime. To widen the accessibility and integration of ssREAD, we intend to develop an R package within the Bioconductor project and a Python library in the future, enabling users to access all datasets both locally and remotely through ssREAD's server-side API. We plan to update the ssREAD database regularly, with new data and features being added every six months. This will ensure that our users have access to the most recent and relevant data in the field of single-cell research. Recognizing the need for a platform that accommodates the diverse range of interests in the scientific community, we plan to include a wider range of sc/snRNA-seq and spatial omics data of neural systems such as human iPSC-derived neurons, glia, and neural organoids. Furthermore, we aspire to include more

neurodegenerative diseases like Frontotemporal lobar degeneration, Parkinson's disease, and Amyotrophic lateral sclerosis, serving a broader research community. Future plans also include expanding our analytical pipelines and visualization methodologies to enrich the capabilities of ssREAD.

## Methods

*ScRNA-seq and snRNA-seq Data Categorization*

In constructing our comprehensive atlas, we meticulously curated a collection of 67 studies that encompass a diverse array of brain regions, spanning both sexes and a wide age range in human and murine models. This extensive compilation yielded a dataset that encapsulates a rich tapestry of cellular profiles. We have methodically reclassified the original datasets into distinct subsets, each meticulously delineated by species (human or mouse), gender (male or female), brain region (including but not limited to the cortex, Middle temporal gyrus, superior frontal gyrus, cerebellum, subventricular zone, superior parietal lobe, and hippocampus), pathological status (disease or control), and age bracket (7, 15, or 20 months for mice; 50–100+ years for humans) (**Supplementary Data 18**).

*ST Data Preprocessing*

Spatial transcriptomics data preprocessing was performed using the Seurat v5 and Squidpy v1.3.0 packages. Seurat was employed for quality control, normalization, and identification of highly variable genes[54]. Squidpy was used for spatially-resolved computations, such as spatial autocorrelation analysis, extracting neighborhood information, and spatially-resolved clustering[55].

*Spatially Variable Genes (SVGs) identification*

To identify spatially variable genes, the spaGCN v1.2.7 was utilized[13]. It integrates graph convolutional network (GCN) with spatial transcriptomics to detect genes that vary significantly across different spatial locations in tissue samples.

*Cell type annotation*

We applied a similar cell type annotation strategy used in scREAD. Our quest for analytical precision led us to modify our initial cell type annotation strategy to increase annotation accuracy. We recognized that several marker genes used in the original release demonstrated sub-optimal specificity. To address this, we executed two iterations of cell type annotation, which involved filtering out several less specific markers. During the first iteration, only Neurons were annotated. Following this, the Neurons were then isolated from the complete dataset, and second iteration marker genes were employed to annotate Excitatory neurons and Inhibitory neurons. The cell labels for Neurons were subsequently replaced by those of Excitatory neurons or Inhibitory neurons. This revised workflow enhances the annotation quality by accounting for the discrepancies between

different cell types and subtypes. The marker gene list can be accessed in **Supplementary Data 19**.

We used SCINA R package that leverages prior marker genes information and simultaneously performs cell type clustering and assignment for known cell types[56]. Furthermore, SCINA shows good performances among prior-knowledge classifiers when high-quality marker genes are provided[57]. Each cell was assigned a cell type based on a manually created marker gene list file using SCINA v1.2.0, whereas the cells with unknown labels marked by SCINA were first compared with predicted clusters from Seurat, and then the unknown labels were assigned to the most dominant cell types within the predicted clusters.

*Cell Type Deconvolution Analysis*

Deconvolution of mixed cell populations was performed using the CARD v1.1 R package[44], which employs a reference-based approach to estimate the proportions of different cell types in bulk RNA-seq or spatial transcriptomics data. It uses cell-type-specific gene expression signatures from a reference dataset to calculate cell type proportions.

*Differentially Expressed Gene and Functional Enrichment Analysis*

DEG analysis was performed using the MAST v 1.28.0 R package[58] through Seurat's FindMarkers function. This function detects DEGs between various conditions, such as disease versus control or male versus female. For overall AD versus control comparisons, age at death and sex were included as covariates. For sex-specific differences in AD versus control comparisons, age at death was included as the covariate. The *p*-values obtained from each comparison were adjusted to control the false discovery rate (FDR) using the Bonferroni correction for all genes in the dataset. Genes with an FDR <0.05 were considered differentially expressed. The enrichment analysis was conducted using Enrichr's API services to identify relevant Gene Ontology (GO) terms and Reactome pathway databases [59,60]. The statistical significance of the enrichment was determined using a hypergeometric test which is a binomial proportion test that assumes a binomial distribution and independence for the probability of any gene belonging to any sets. The *p*-values obtained from Enrichr were adjusted to *q*-values using the Benjamini-Hochberg correction method.

*Cell-cell communication analysis*

Cellular communication was investigated using the CellChat v2[61] R package toolkit for the inference, visualization, and analysis of cell-cell communication patterns from single-cell and spatial transcriptomics data. CellChat v2 extends to analyze communication patterns among neighboring cell clusters within spatially resolved

transcriptomic landscapes, with an expanded communication network interactions database encompassing a comprehensive compendium of over 1,000 interactions.

*Tissue Architecture Identification*

The tissue architecture was analyzed using RESEPT v1[15], a computational tool that generates an atlas of regional gene expression patterns in spatial transcriptomics data. This tool enables the visualization of how gene expression varies across different spatial regions within a tissue sample.

*Gene Regulatory Network Analysis*

Regulatory analysis was conducted using DeepMAPS v1.0[62,63]. We employed DeepMAPS to build the cell and gene embeddings and obtained active gene modules through the graph transformer model. The gene modules were further sent to DeepMAPS to perform cell cluster active gene module determination, de novo motif finding, and TF matching and CTSR determination with the default parameters. Gene regulatory networks were constructed to indicate the predicted TF-gene regulatory relations via Cytoscape[64,65].

*User interface*

We made substantial improvements to the user interface to enhance accessibility and ease of use. Key modifications were made within the dataset details page, such as the addition of gene expression display for single cells/spots using feature plots and violin plots. We also introduced a dedicated query page for DEG searching. Notably, search results can now be refined based on sex, group, and condition parameters. The new design offers clearer demarcation between DEG searching and query options. Moreover, we integrated the function to calculate overlapping DEGs from multiple comparisons, an approach originally outlined in the scREAD protocol[66]. This functionality is now delivered through an interactive online query. We also redesigned the homepage for better usability, positioning the search bar at the top and displaying key information about the number of species, studies, assays, and versions for higher visibility.

*ssREAD server construction*

ssREAD is hosted on an HPE XL675d RHEL system outfitted with a 2×128-core AMD EPYC 7H12 CPU, 64GB RAM, and 8×NVIDIA A100 80GB GPU. Our backend server, written in TypeScript and built with the koa.js framework, leverages Auth0 to provide independent user authentication and authorization services. Our frontend is constructed with NUXT, utilizing Vuetify as the UI library and Plotly.js for data visualization. Communication between frontend and backend servers is enabled using REST API. This streamlined server construction ensures robust, efficient, and scalable performance of ssREAD.

## Data availability

All data used were sourced from public collections, and are detailed with online accession numbers in Supplementary Data 1. ssREAD is freely available at https://bmblx.bmi.osumc.edu/ssread/. A backup link is also provided at https://go.osu.edu/ssread. Relevant raw data from each figure is available in the Source Data file. The processed data in this study can be downloaded through the link https://bmblx.bmi.osumc.edu/ssread/downloads.

## Code availability

The frontend code is available at https://github.com/OSU-BMBL/ssread. The backend code is available at https://github.com/OSU-BMBL/ssread-backend. Additionally, a real-time ssREAD server status page is available at https://ssread.statuspage.io/. The source code to the version of the code used in this study is also available on Zenodo[67].

## Reference


1 Nichols, E. *et al.* Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health* **7**, e105-e125 (2022). https://doi.org:10.1016/S2468-2667(21)00249-8

2 Wu, Y.-T. *et al.* The changing prevalence and incidence of dementia over time - current evidence. *Nature Reviews. Neurology* **13**, 327-339 (2017). https://doi.org:10.1038/nrneurol.2017.63

3 Jiang, J., Wang, C., Qi, R., Fu, H. & Ma, Q. scREAD: A Single-Cell RNA-Seq Database for Alzheimer's Disease. *iScience* **23**, 101769 (2020). https://doi.org:10.1016/j.isci.2020.101769

4 Marshall, C. R., Farrow, M. A., Djambazova, K. V. & Spraggins, J. M. Untangling Alzheimer's disease with spatial multi-omics: a brief review. *Front Aging Neurosci* **15**, 1150512 (2023). https://doi.org:10.3389/fnagi.2023.1150512

5 Zhou, Y. *et al.* The Alzheimer's Cell Atlas (TACA): A single-cell molecular map for translational therapeutics accelerator in Alzheimer's disease. *Alzheimers Dement (N Y)* **8**, e12350 (2022). https://doi.org:10.1002/trc2.12350

6 Zhao, T. *et al.* SC2disease: a manually curated database of single-cell transcriptome for human diseases. *Nucleic Acids Research* **49**, D1413-D1419 (2020). https://doi.org:10.1093/nar/gkaa838

7 Li, X. W. *et al.* SCAD-Brain: a public database of single cell RNA-seq data in human and mouse brains with Alzheimer's disease. *Front Aging Neurosci* **15**, 1157792 (2023). https://doi.org:10.3389/fnagi.2023.1157792

8 Guo, T. *et al.* Molecular and cellular mechanisms underlying the pathogenesis of Alzheimer’s disease. *Molecular Neurodegeneration* **15**, 40 (2020). https://doi.org:10.1186/s13024-020-00391-7

9 Chen, S. *et al.* Spatially resolved transcriptomics reveals genes associated with the vulnerability of middle temporal gyrus in Alzheimer’s disease. *Acta Neuropathologica Communications* **10**, 188 (2022). https://doi.org:10.1186/s40478-022-01494-6

10 Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research* **41**, D991-D995 (2013). https://doi.org:10.1093/nar/gks1193

11 Yuan, Z. *et al.* SODB facilitates comprehensive exploration of spatial omics data. *Nat Methods* **20**, 387-399 (2023). https://doi.org:10.1038/s41592-023-01773-7

12 Xu, Z. *et al.* STOmicsDB: a database of Spatial Transcriptomic data. *bioRxiv*, 2022.2003.2011.481421 (2022). https://doi.org:10.1101/2022.03.11.481421

13 Hu, J. *et al.* SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature Methods* **18**, 1342-1351 (2021). https://doi.org:10.1038/s41592-021-01255-8

14 Chen, S. *et al.* Wolframin is a novel regulator of tau pathology and neurodegeneration. *Acta Neuropathologica* (2022). https://doi.org:10.1007/s00401-022-02417-4

15 Chang, Y. *et al.* Define and visualize pathological architectures of human tissues from spatially resolved transcriptomics using deep learning. *Computational and Structural Biotechnology Journal* **20**, 4600-4617 (2022). https://doi.org:10.1016/j.csbj.2022.08.029

16 Arnold, S. E., Hyman, B. T., Flory, J., Damasio, A. R. & Van Hoesen, G. W. The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. *Cereb Cortex* **1**, 103-116 (1991). https://doi.org:10.1093/cercor/1.1.103

17 Braak, H. & Braak, E. On areas of transition between entorhinal allocortex and temporal isocortex in the human brain. Normal morphology and lamina-specific pathology in Alzheimer's disease. *Acta Neuropathol* **68**, 325-332 (1985). https://doi.org:10.1007/bf00690836

18 Morrison, J. H. & Hof, P. R. Selective vulnerability of corticocortical and hippocampal circuits in aging and Alzheimer's disease. *Prog Brain Res* **136**, 467-486 (2002). https://doi.org:10.1016/s0079-6123(02)36039-4

19 Allen, C., Chang, Y., Ma, Q. & Chung, D. MAPLE: A Hybrid Framework for Multi-Sample Spatial Transcriptomics Data. *bioRxiv*, 2022.2002.2028.482296 (2022). https://doi.org:10.1101/2022.02.28.482296

20 Sweeney, G. & Song, J. The association between PGC-1α and Alzheimer's disease. *Anat Cell Biol* **49**, 1-6 (2016). https://doi.org:10.5115/acb.2016.49.1.1

21 Sheng, B. *et al.* Impaired mitochondrial biogenesis contributes to mitochondrial dysfunction in Alzheimer's disease. *J Neurochem* **120**, 419-429 (2012). https://doi.org:10.1111/j.1471-4159.2011.07581.x

22 Swerdlow, R. H., Burns, J. M. & Khan, S. M. The Alzheimer's disease mitochondrial cascade hypothesis: progress and perspectives. *Biochim Biophys Acta* **1842**, 1219-1231 (2014). https://doi.org:10.1016/j.bbadis.2013.09.010

23 Bennett, J. P., Jr. & Keeney, P. M. Alzheimer's and Parkinson's brain tissues have reduced expression of genes for mtDNA OXPHOS Proteins, mitobiogenesis regulator PGC-1α protein and mtRNA stabilizing protein LRPPRC (LRP130). *Mitochondrion* **53**, 154-157 (2020). https://doi.org:10.1016/j.mito.2020.05.012

24 Hou, J., Chen, Y., Grajales-Reyes, G. & Colonna, M. TREM2 dependent and independent functions of microglia in Alzheimer's disease. *Mol Neurodegener* **17**, 84 (2022). https://doi.org:10.1186/s13024-022-00588-y

25 Chen, X. & Holtzman, D. M. Emerging roles of innate and adaptive immunity in Alzheimer's disease. *Immunity* **55**, 2236-2254 (2022). https://doi.org:10.1016/j.immuni.2022.10.016

26 Salter, M. W. & Stevens, B. Microglia emerge as central players in brain disease. *Nat Med* **23**, 1018-1027 (2017). https://doi.org:10.1038/nm.4397

27 Gate, D. *et al.* Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. *Nature* **577**, 399-404 (2020). https://doi.org:10.1038/s41586-019-1895-7

28 Sun, N. *et al.* Human microglial state dynamics in Alzheimer's disease progression. *Cell* **186**, 4386-4403.e4329 (2023). https://doi.org:10.1016/j.cell.2023.08.037

29 Sala Frigerio, C. *et al.* The Major Risk Factors for Alzheimer's Disease: Age, Sex, and Genes Modulate the Microglia Response to Aβ Plaques. *Cell Rep* **27**, 1293-1306.e1296 (2019). https://doi.org:10.1016/j.celrep.2019.03.099

30 Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290.e1217 (2017). https://doi.org:10.1016/j.cell.2017.05.018

31 Zhou, X. *et al.* Molecular differences in brain regional vulnerability to aging between males and females. *Front Aging Neurosci* **15**, 1153251 (2023). https://doi.org:10.3389/fnagi.2023.1153251

32 Tramutola, A., Di Domenico, F., Barone, E., Perluigi, M. & Butterfield, D. A. It Is All about (U)biquitin: Role of Altered Ubiquitin-Proteasome System and UCHL1 in Alzheimer Disease. *Oxid Med Cell Longev* **2016**, 2756068 (2016). https://doi.org:10.1155/2016/2756068

33 Ma, A. *et al.* Single-cell biological network inference using a heterogeneous graph transformer. *Nature Communications* **14** (2023). https://doi.org:10.1038/s41467-023-36559-0

34 Du, Y. *et al.* Activating transcription factor 6 reduces Aβ1-42 and restores memory in Alzheimer's disease model mice. *Int J Neurosci* **130**, 1015-1023 (2020). https://doi.org:10.1080/00207454.2020.1715977

35 Hu, Y. T. *et al.* Early growth response-1 regulates acetylcholinesterase and its relation with the course of Alzheimer's disease. *Brain Pathol* **29**, 502-512 (2019). https://doi.org:10.1111/bpa.12688

36 Qin, X., Wang, Y. & Paudel, H. K. Inhibition of Early Growth Response 1 in the Hippocampus Alleviates Neuropathology and Improves Cognition in an Alzheimer Model with Plaques and Tangles. *Am J Pathol* **187**, 1828-1847 (2017). https://doi.org:10.1016/j.ajpath.2017.04.018

37 Miller, J. A. *et al.* SEA-AD: Scientific analysis and open access resources targeting early changes in Alzheimer's disease. *Alzheimer's & Dementia* **19**, e063478 (2023). https://doi.org:https://doi.org/10.1002/alz.063478

38 Krasemann, S. *et al.* The TREM2-APOE Pathway Drives the Transcriptional Phenotype of Dysfunctional Microglia in Neurodegenerative Diseases. *Immunity* **47**, 566-581.e569 (2017). https://doi.org:10.1016/j.immuni.2017.08.008

39 Mathys, H. *et al.* Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365-4385.e4327 (2023). https://doi.org:10.1016/j.cell.2023.08.039

40 Rivera-Escalera, F. *et al.* IL-1β-driven amyloid plaque clearance is associated with an expansion of transcriptionally reprogrammed microglia. *J Neuroinflammation* **16**, 261 (2019). https://doi.org:10.1186/s12974-019-1645-7

41 Swaminathan, S. *et al.* Analysis of copy number variation in Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. *PLoS One* **7**, e50640 (2012). https://doi.org:10.1371/journal.pone.0050640

42 McMillan, P. J. *et al.* Pathological tau drives ectopic nuclear speckle scaffold protein SRRM2 accumulation in neuron cytoplasm in Alzheimer's disease. *Acta Neuropathol Commun* **9**, 117 (2021). https://doi.org:10.1186/s40478-021-01219-1

43 Lester, E. *et al.* Cytosolic condensates rich in polyserine define subcellular sites of tau aggregation. *Proc Natl Acad Sci U S A* **120**, e2217759120 (2023). https://doi.org:10.1073/pnas.2217759120

44 Ma, Y. & Zhou, X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nature Biotechnology* **40**, 1349-1359 (2022). https://doi.org:10.1038/s41587-022-01273-7

45 Kajiwara, Y. *et al.* GJA1 (connexin43) is a key regulator of Alzheimer's disease pathogenesis. *Acta Neuropathologica Communications* **6**, 144 (2018). https://doi.org:10.1186/s40478-018-0642-x

46 Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019). https://doi.org:10.1038/s41586-019-1195-2

47 Yang, A. C. *et al.* A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885-892 (2022). https://doi.org:10.1038/s41586-021-04369-3

48 Coales, I. *et al.* Alzheimer's disease-related transcriptional sex differences in myeloid cells. *J Neuroinflammation* **19**, 247 (2022). https://doi.org:10.1186/s12974-022-02604-w

49 Alquraini, A. & El Khoury, J. Scavenger receptors. *Curr Biol* **30**, R790-r795 (2020). https://doi.org:10.1016/j.cub.2020.05.051

50 Wilkinson, K. & El Khoury, J. Microglial scavenger receptors and their roles in the pathogenesis of Alzheimer's disease. *Int J Alzheimers Dis* **2012**, 489456 (2012). https://doi.org:10.1155/2012/489456

51 Edler, M. K., Johnson, C. T., Ahmed, H. S. & Richardson, J. R. Age, sex, and regional differences in scavenger receptor CD36 in the mouse brain: Potential relevance to cerebral amyloid angiopathy and Alzheimer's disease. *Journal of Comparative Neurology* **529**, 2209-2226 (2021). https://doi.org:https://doi.org/10.1002/cne.25089

52 Butovsky, O. & Weiner, H. L. Microglial signatures and their role in health and disease. *Nat Rev Neurosci* **19**, 622-635 (2018). https://doi.org:10.1038/s41583-018-0057-5

53 Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019). https://doi.org:10.1038/s41586-019-1195-2

54 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411-420 (2018). https://doi.org:10.1038/nbt.4096

55 Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nature Methods* **19**, 171-178 (2022). https://doi.org:10.1038/s41592-021-01358-2

56 Zhang, Z. *et al.* SCINA: A Semi-Supervised Subtyping Algorithm of Single Cells and Bulk Samples. *Genes* **10**, 531 (2019). https://doi.org:10.3390/genes10070531

57 Abdelaal, T. *et al.* A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biology* **20**, 194 (2019). https://doi.org:10.1186/s13059-019-1795-z

58 Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biology* **16**, 278 (2015). https://doi.org:10.1186/s13059-015-0844-5

59 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research* **44**, W90-W97 (2016). https://doi.org:10.1093/nar/gkw377

60 Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Research* **50**, D687-D692 (2021). https://doi.org:10.1093/nar/gkab1028

61 Jin, S., Plikus, M. V. & Nie, Q. CellChat for systematic analysis of cell-cell communication from single-cell and spatially resolved transcriptomics. *bioRxiv*, 2023.2011.2005.565674 (2023). https://doi.org:10.1101/2023.11.05.565674

62 Ma, A. *et al.* IRIS3: integrated cell-type-specific regulon inference server from single-cell RNA-Seq. *Nucleic Acids Research* **48**, W275-W286 (2020). https://doi.org:10.1093/nar/gkaa394

63 Ma, A. *et al.* Single-cell biological network inference using a heterogeneous graph transformer. *Nat Commun* **14**, 964 (2023). https://doi.org:10.1038/s41467-023-36559-0

64 Gustavsen, J. A., Pai, S., Isserlin, R., Demchak, B. & Pico, A. R. RCy3: Network biology using Cytoscape from within R. Report No. 8:1774, (F1000Research, 2019).
65 Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**, 2498-2504 (2003). https://doi.org:10.1101/gr.1239303
66 Wang, C., Xiang, Y., Fu, H. & Ma, Q. Use of scREAD to explore and analyze single-cell and single-nucleus RNA-seq data for Alzheimer's disease. *STAR Protocols* **2**, 100513 (2021). https://doi.org:10.1016/j.xpro.2021.100513
67 Wang, C. A single-cell and spatial RNA-seq database for Alzheimer's Disease. *Zenodo* (2024). https://doi.org:10.5281/zenodo.10992583

**Acknowledgments**

This work was supported by awards R21HG012482 (Q.M.) and R01AG075092 (H.F.) from the National Institutes of Health (NIH). We would like to acknowledge Mr. Qi Guo, Mr. Shaohong Feng for their kind help in testing the website and database. We also thank Mr. Nicholas Sweeney and Ms. Tae Yeon Kim for their helpful discussions during the revision.

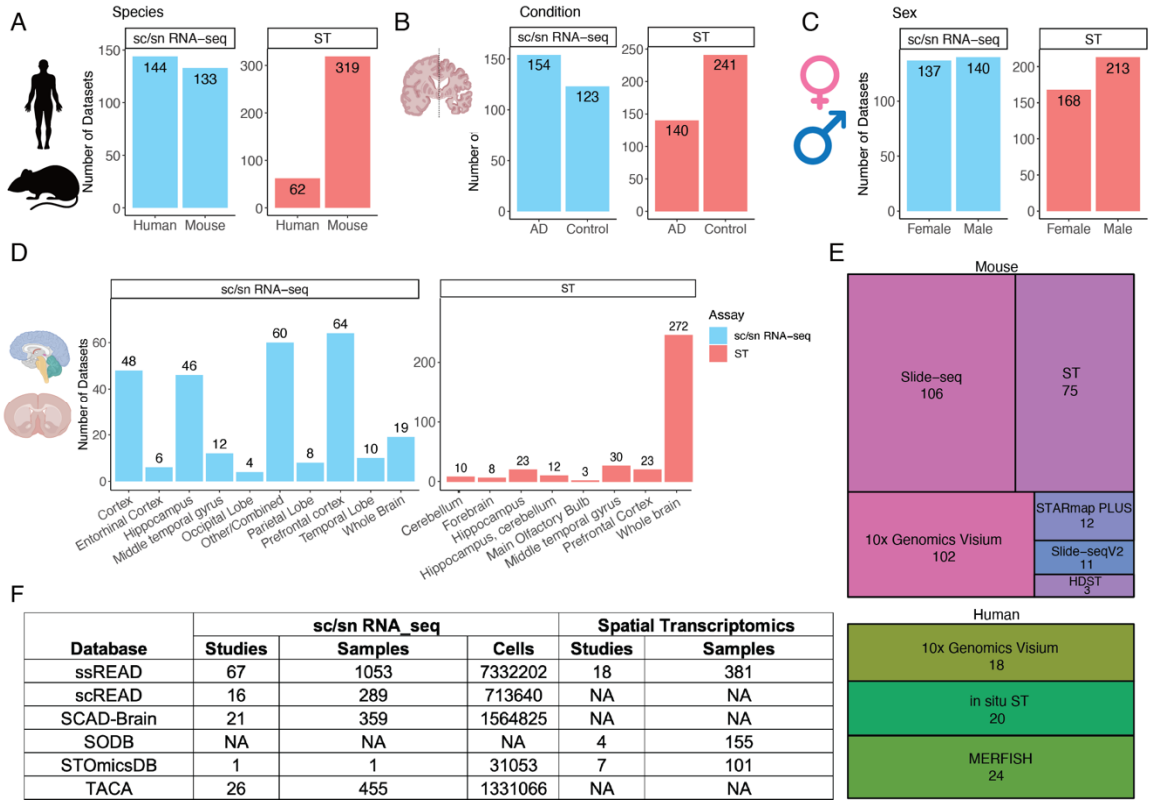
**Author Contributions Statement**

Q.M. and H.F. designed the manuscript contents and experiments. C.W. and M.M. contributed to data collection and analysis. C.W. developed and implemented the database and the API. D.A., M.M., and A.M. tested the database and interpreted the results. C.W., D.A., A.M., and J.B. drafted the manuscript. All authors revised the final manuscript.

**Competing Interests Statement**

The authors declare no competing interests.

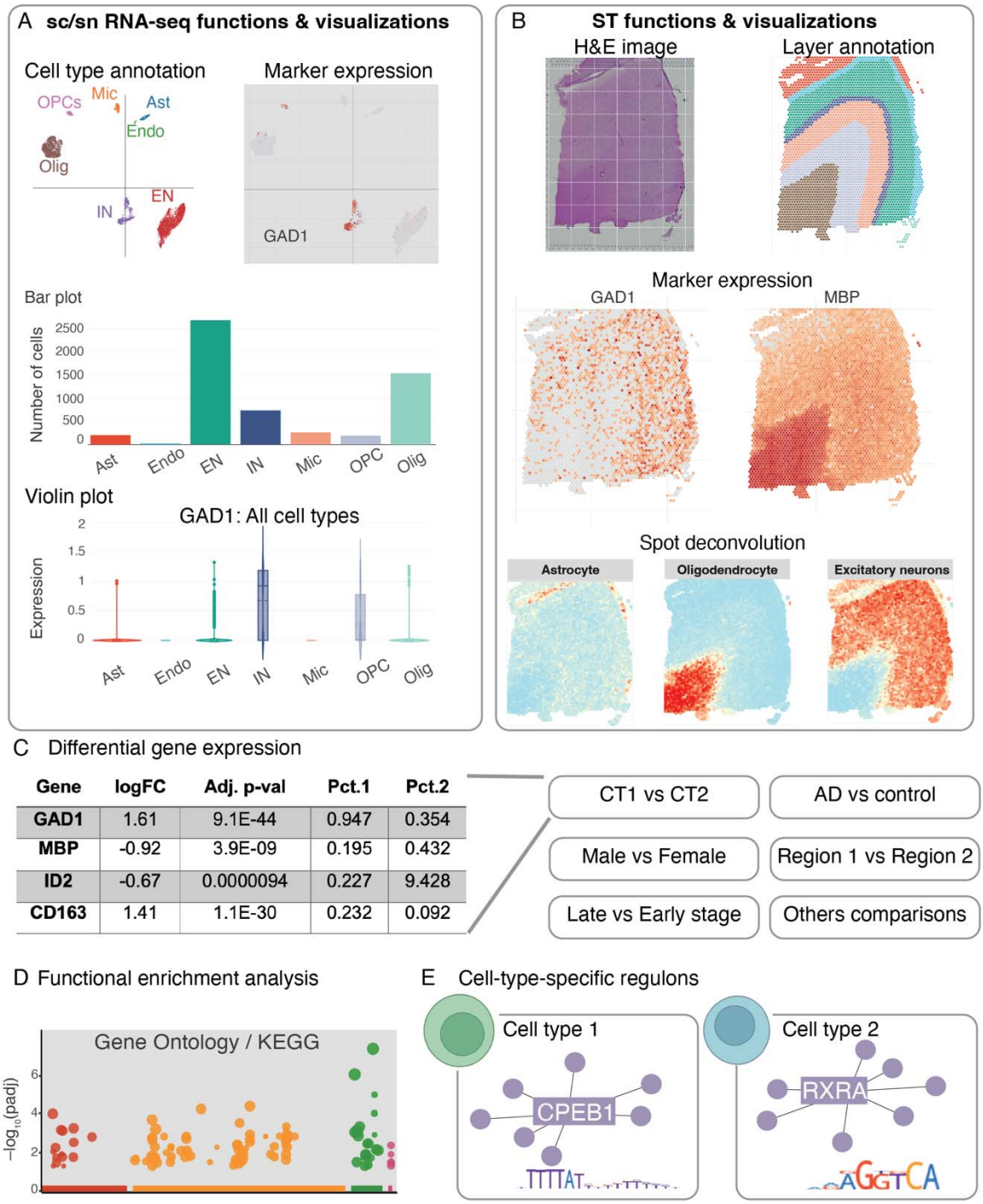
**Figures**




F

| Database | sc/sn RNA_seq | | | Spatial Transcriptomics | |
|---|---|---|---|---|---|
| | Studies | Samples | Cells | Studies | Samples |
| ssREAD | 67 | 1053 | 7332202 | 18 | 381 |
| scREAD | 16 | 289 | 713640 | NA | NA |
| SCAD-Brain | 21 | 359 | 1564825 | NA | NA |
| SODB | NA | NA | NA | 4 | 155 |
| STOmicsDB | 1 | 1 | 31053 | 7 | 101 |
| TACA | 26 | 455 | 1331066 | NA | NA |

**Fig. 1.** ssREAD Data characteristics and statistics (as of December 2023). (A-D) Barplots show the number of datasets by species, sex, condition, and brain regions, respectively. Figure created with BioRender.com. (E) Treemap presents the breakdown of technologies deployed in mouse and human studies. (F) The comparative table showcases the number of studies and datasets across extant AD databases and data collection sources.




| Gene | logFC | Adj. p-val | Pct.1 | Pct.2 |
|---|---|---|---|---|
| **GAD1** | 1.61 | 9.1E-44 | 0.947 | 0.354 |
| **MBP** | -0.92 | 3.9E-09 | 0.195 | 0.432 |
| **ID2** | -0.67 | 0.0000094 | 0.227 | 9.428 |
| **CD163** | 1.41 | 1.1E-30 | 0.232 | 0.092 |

**Fig. 2**. Overview of ssREAD functions. (A) Features and visual representations related to sc/snRNA-seq, encompassing cell type annotations, marker gene expressions, and graphic depictions via bar and violin plots. Each box showcases the minimum, first quartile, median, third quartile, and maximum average expression values of a cell type (Ast: n□=□208, Endo: n□=□28, EN: n□=□2,696, IN: n□=□740, MIC: n□=□266, OPC: n□=□195, and Olig: n□=□1,546). Dots represent outliers. (B) Functions and visualizations pertinent to ST, highlighting H&E imagery, layer annotations, marker gene

expressions, and spot deconvolutions. (C) DEG analysis, with comparisons drawn between categories like Cell Type 1 vs. Cell Type 2 (CT1 vs. CT2), AD vs. Control, Male vs. Female, Brain Regions 1 vs. 2, etc. *p*-values were calculated based on two-sided Wilcoxon Rank-Sum test and adjusted using Bonferroni correction. (D) Functional enrichment analysis focusing on GO ontology and KEGG pathways. *p*-values were calculated using the Hypergeometric test from Enrichr and were adjusted using the Benjamini-Hochberg correction method. (E) Predictions of cell type-specific regulons. The following abbreviations are used for cell types: Astrocytes (Ast), Endothelial Cells (Endo), Excitatory Neurons (EN), Inhibitory Neurons (IN), Microglia (Mic), Oligodendrocytes (Olig), and Oligodendrocyte Precursor Cells (OPC).

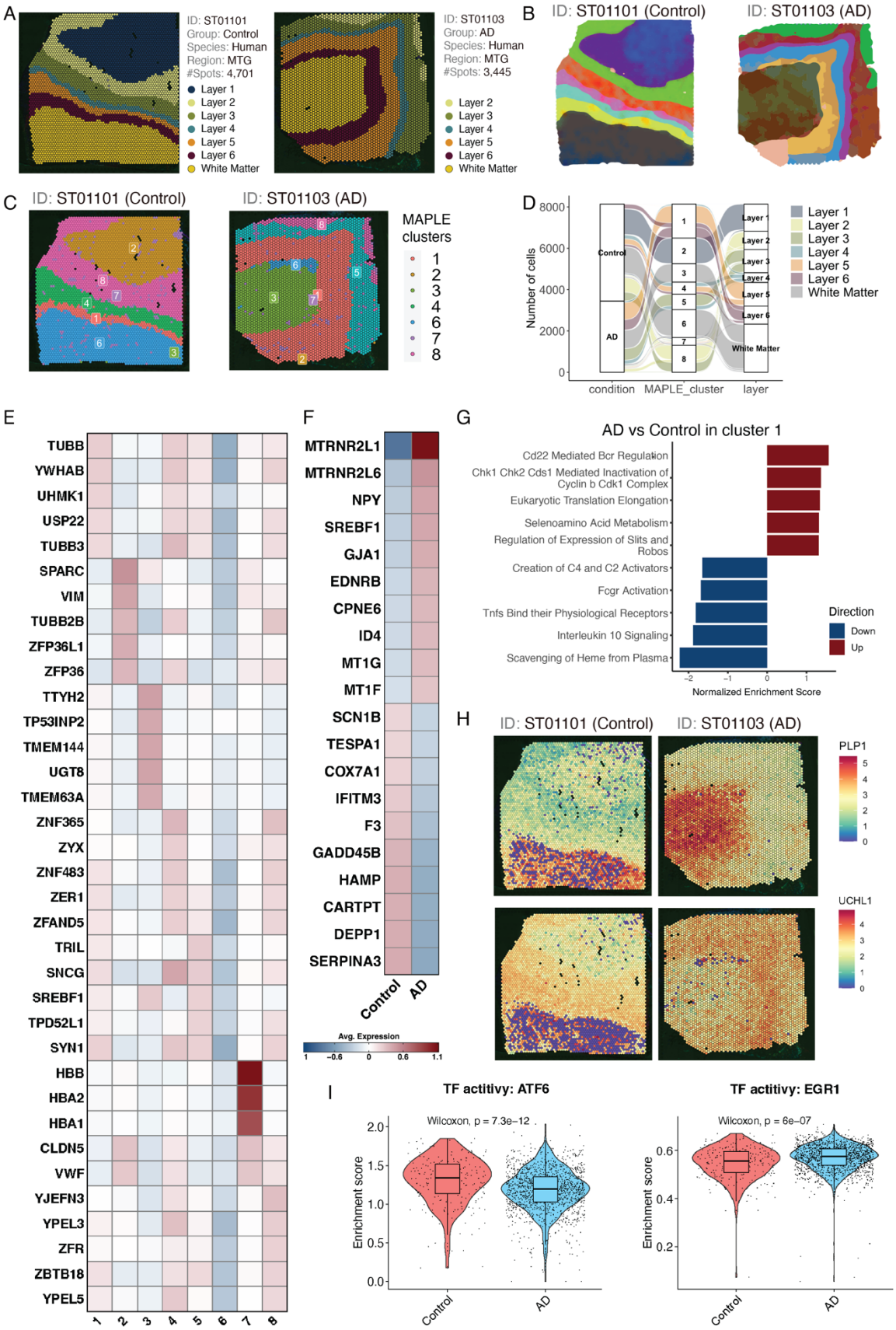

A
ID: ST01101
Group: Control
Species: Human
Region: MTG
#Spots: 4,701
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
Layer 6
White Matter
ID: ST01103
Group: AD
Species: Human
Region: MTG
#Spots: 3,445
Layer 2
Layer 3
Layer 4
Layer 5
Layer 6
White Matter
B
ID: ST01101 (Control)
ID: ST01103 (AD)
C
ID: ST01101 (Control)
ID: ST01103 (AD)
MAPLE clusters
1
2
3
4
6
7
8
D
Number of cells
Control
AD
condition
MAPLE_cluster
layer
Layer 1
Layer 2
Layer 3
Layer 4
Layer 5
Layer 6
White Matter
E
TUBB
YWHAB
UHMK1
USP22
TUBB3
SPARC
VIM
TUBB2B
ZFP36L1
ZFP36
TTYH2
TP53INP2
TMEM144
UGT8
TMEM63A
ZNF365
ZYX
ZNF483
ZER1
ZFAND5
TRIL
SNCG
SREBF1
TPD52L1
SYN1
HBB
HBA2
HBA1
CLDN5
VWF
YJEFN3
YPEL3
ZFR
ZBTB18
YPEL5
F
MTRNR2L1
MTRNR2L6
NPY
SREBF1
GJA1
EDNRB
CPNE6
ID4
MT1G
MT1F
SCN1B
TESPA1
COX7A1
IFITM3
F3
GADD45B
HAMP
CARTPT
DEPP1
SERPINA3
Control
AD
Avg. Expression
G
AD vs Control in cluster 1
Cd22 Mediated Bcr Regulation
Chk1 Chk2 Cds1 Mediated Inactivation of Cyclin b Cdk1 Complex
Eukaryotic Translation Elongation
Selenoamino Acid Metabolism
Regulation of Expression of Slits and Robos
Creation of C4 and C2 Activators
Fcgr Activation
Tnfs Bind their Physiological Receptors
Interleukin 10 Signaling
Scavenging of Heme from Plasma
Direction
Down
Up
Normalized Enrichment Score
H
ID: ST01101 (Control)
ID: ST01103 (AD)
PLP1
UCHL1
I
TF actitivy: ATF6
Wilcoxon, p = 7.3e−12
Enrichment score
Control
AD
TF actitivy: EGR1
Wilcoxon, p = 6e−07
Enrichment score
Control
AD

**Fig. 3**. Multi-dimensional analysis of spatially-informed sub-populations. (A) Annotation of the six cortical layers alongside the adjacent white matter within two human middle temporal gyrus (MTG) brain samples (ST01101 and ST01103). (B) Detection of spatial domains by RESEPT. (C) Visualization using MAPLE to elucidate shared or unique spatial domains identified across the two Spatial Transcriptomics samples (ST01101 and ST01103). (D) Alluvial diagrams showcasing the progression of cells: originating from individual samples, aggregating into joint subpopulations, and culminating in layer annotations. (E) A heatmap depicting genes specific to the MAPLE-derived clusters for both AD and Control samples. (F) Heatmap representing the top 10 upregulated and top 10 downregulated genes distinguishing AD from Control within Cluster 1. (G) Gene Set Enrichment Analysis (GSEA) of DEGs from (E) plotted against REACTOME pathways. The bar plot shows the top 10 upregulated and downregulated pathways, accompanied by normalized enrichment scores. (H) Spatial feature plots highlighting the variance in gene expression of *PLP1* and *UCHL1* from Cluster 1, segregated by AD and Control samples. (I) Violin plots showcasing the activity of two selected TFs between AD and Control, with associated *p*-values calculated from a two-sided Wilcoxon rank-sum test. Each box showcases the minimum, first quartile, median, third quartile, and maximum ARI results of a tool performed on different data subsets (Control group: n = 232, and AD group: n = 1,412). Dots represent spatial spots. Source data are provided as a Source Data file.

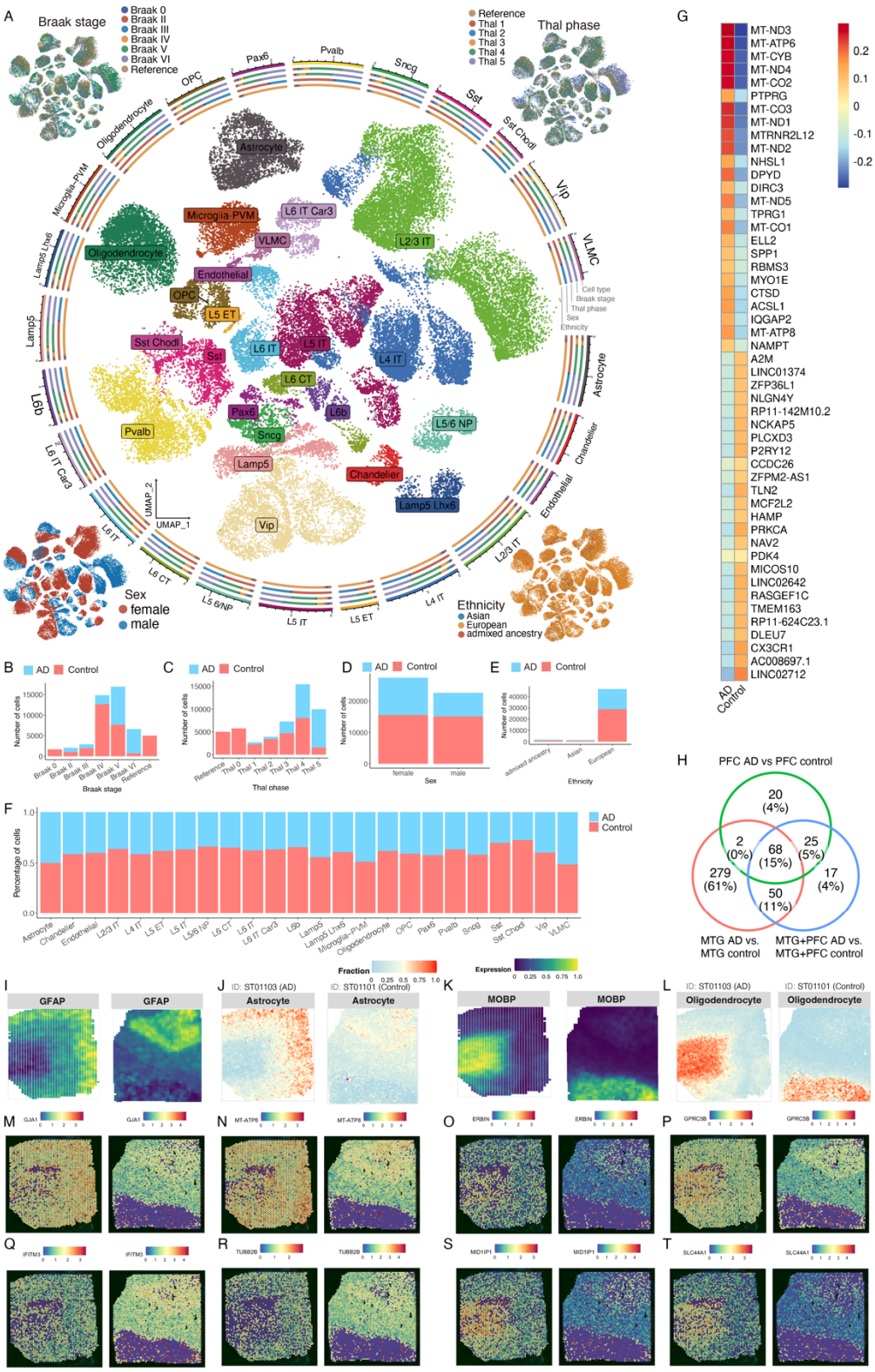



**Fig. 4.** ssREAD Facilitates Comprehensive Spatial Transcriptomics Analyses Synergized with scRNA-seq Datasets. (A) UMAP representation of cell types derived from the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD). Clockwise from the top left, corner insets

elucidate the Braak stage, Thal phase, ethnicity, and sex attributes. (B-E) Bar plots visualizing the distribution of cells based on the Braak stage, Thal phase, sex, and ethnicity, categorized by the condition in the atlas. (F) Bar plots displaying the fractional representation of cell types, contrasting AD and control within the SEA-AD atlas. (G) Heatmap of top and bottom 25 DEGs identified between AD and control samples. (H) Comparison of DEGs identified in integrated dataset to additional data sets. (I-J) Deconvolution analysis of cell types between ST samples (ST01101 and ST01103) and the scRNA-seq SEA-AD atlas, presenting the cell type fractions for oligodendrocytes, astrocytes, and endothelial cells. The cell type fractions are positioned beneath their corresponding marker gene expression indicators: *MOBP* for oligodendrocytes, *GFAP* for astrocytes, and *CLDN5* for endothelial cells, with analyses segregated into control (I) and AD (J) conditions.

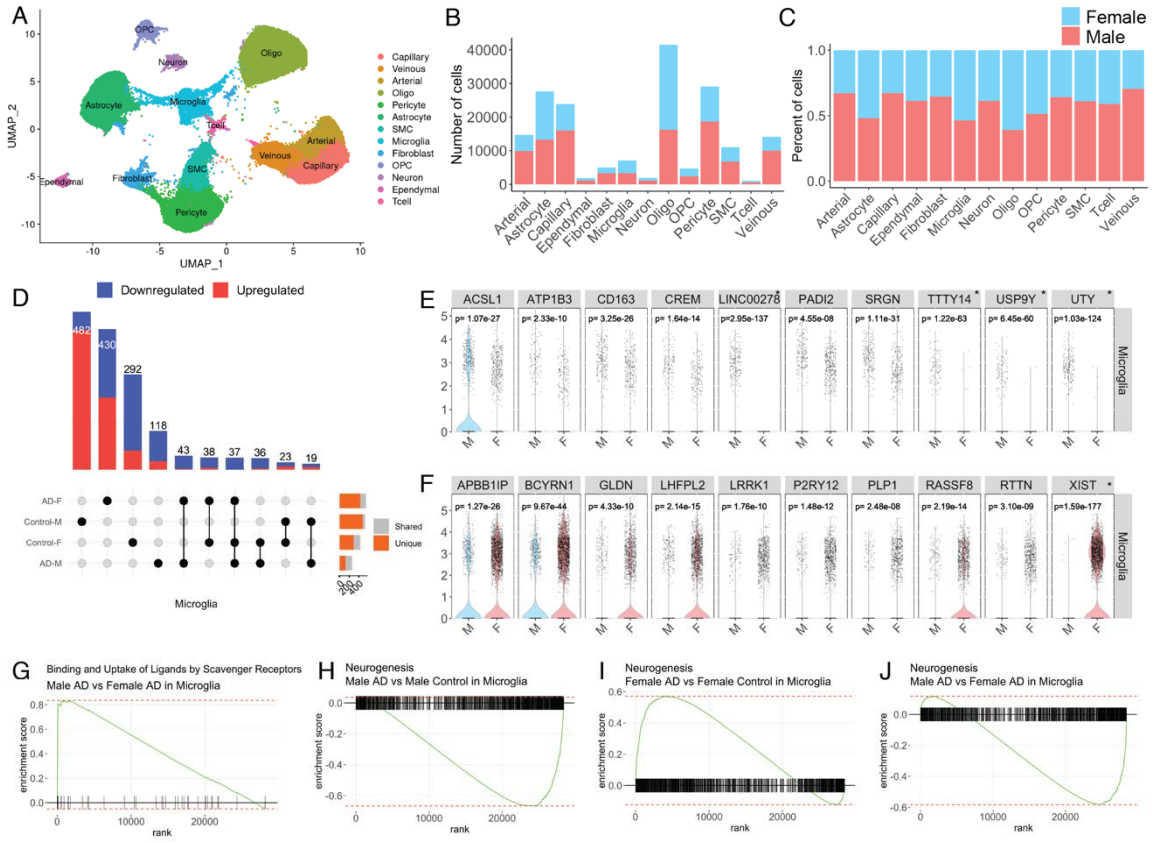



**Fig. 5**. Exploration of Sex-Specific Differences at the Cellular Level in AD (A) UMAP visualization of the single-cell data used in the analysis, with different cell types color-coded. (B-C) Bar plots illustrate the count and proportion of each cell type, segregated by sex. This reveals any potential differences in cellular composition between male and female samples. (D) UpSet plot showing the unique and shared DEGs across four groups: Male AD patients, Female AD patients, Male controls, and Female controls. (E) Violin plots for the top 10 upregulated DEGs between male and female in microglia. *

Indicates sex-chromosomal genes. *p*-values were calculated based on a two-sided Wilcoxon Rank-Sum test and adjusted using Bonferroni correction (F) Violin plots for the top 10 downregulated DEGs between male and female in microglia. *p*-values were calculated based on a two-sided Wilcoxon Rank-Sum test and adjusted using Bonferroni correction. (G) Gene Set Enrichment Analysis (GSEA) plot showing the enrichment of genes involved in binding and uptake of ligands by scavenger receptors. (H-J) GSEA plots showing the enrichment of genes involved in neurogenesis for three different comparisons: Male AD vs. Male Control in Microglia (H), Female AD vs. Female Control in Microglia (I), and Male AD vs. Female AD in Microglia (J). These plots highlight the sex-specific differences in neurogenesis-related gene activity under AD conditions.